

ciency among six distinct hybrid algorithms used with health care data. **METHODS:** Six matching algorithms were examined. Each combined covariate matching with a different propensity scoring function: continuous factor, weighting factor, caliper, parenting factor, nesting factor or partner. The algorithms were compared in terms of 1:1 matching rate, computing time, bias balancing and standardized difference. The influence of sample size variation on stability and efficiency was considered. Paired T-test, Pearson Chi-Square and Standardized Difference were adopted for assessment. **RESULTS:** The superiority of some hybrid algorithms over pure covariate matching was observed. In terms of matching rate, the partner function reported the highest rate (99.7%), followed by its function as a caliper (88.4%), while the parenting function produced the lowest rate (59.5%). All others performed at a similar level. Computing time varied, the most efficient using the propensity score as a parenting factor (00:25:10). The longest reported times were seen when used as a weighting factor (00:37:56) or caliper function (00:37:52). Differences are more profound in large samples. In bias balancing tests, all algorithms were balanced on categorical covariates except when the propensity score was used as a partner or a caliper where each displayed the lowest capability of producing p-values above 0.05. Significant reduction in standardized difference below 10% was indicative of higher efficiency of the hybrid algorithms. Categorical covariates produced values near zero despite the lower performance for the partner approach. With increasing sample size, all investigations performed as expected. **CONCLUSION:** Overall, these hybrid applications exhibited greater efficiency in simultaneously overcoming high dimensionality on covariate matching and reducing variation in propensity score matching. Depending on data characteristics and research profiles, each application has specific merits in certain circumstances.

PMC48

SETTING THE OPTIMAL SCREENING TOOL THRESHOLD FOR A CHRONIC UNDERDIAGNOSED ILLNESS: WHOSE BURDEN MATTERS MOST?

Yu HT, Calimlim B, Dean BB, Dubois R

Cerner LifeSciences, Beverly Hills, CA, USA

Ankylosing spondylitis (AS) is a chronic disease with approximately 0.5% prevalence in the general population and 5% among chronic back pain (CBP) sufferers. The disease typically remains undiagnosed for over a decade which is problematic since new treatments may alter the natural history. An AS screening instrument based upon patient reported data was developed but selecting the optimal screening tool threshold is a critical issue for discussion. Question items were identified from a literature review, patient focus groups, and an advisory board of rheumatologists. A case-control study was conducted to test the screening instrument among subjects with confirmed AS (cases) or CBP for ≥ 3 months (controls). Question items were examined in a multivariate logistic framework using best subsets modeling. Receiver-operator characteristic analysis was conducted to determine optimal sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) of the instrument: AS prevalence set equal to 0.5% in the general population and 5% among CBP sufferers. Responses from 102 cases and 214 controls were analyzed to develop a twelve-variable model. Sensitivities ranging from 69.6% to 90.2% were associated with specificities of 99.1% to 79.9%, respectively. Lowering sensitivities reduced the portion of false positives seen by the provider from 95% to 20.3% (78.6% reduction) and 99.5% to 72.8% (26.8% reduction) for the CBP and general populations, respectively (SE = 69.6%, SP = 99.1%). Selecting the optimal screening tool threshold depends on whose burden matters the most:

increasing sensitivity of the instrument would increase the probability of identifying patients with disease earlier and the ability to improve AS patient well-being. However, this approach would increase the economic burden (additional medical evaluations) from the payer perspective, raise the patient care burden from the rheumatologist perspective, and reduce quality of life for those with false positives. We will discuss the trade-offs in this real world example.

PMC49

TRANSLATING HETEROGENEITY BIAS FROM HEALTH STATUS IN OUTCOMES STUDIES—USING LATENT CLASS CLUSTER ANALYSIS AND LONGITUDINAL DATA

Ahn J

University of Southern California, Los Angeles, CA, USA

Ignoring heterogeneity in health may bias measurement of intervention outcomes through confounding with intervention of interest. If repeated observations on each subject are available, heterogeneity may be usefully included in outcomes studies. We assume heterogeneous health status as a latent index and multiple health proxies (and their correlations) are used to estimate heterogeneous health grouping from the latent index. For example, in a treatment effect study with longitudinal data: 1) estimate K, the number of heterogeneous groups, by latent class cluster analysis (LCCA) using health proxies of each subject at each period, such as comorbidity indices, length of hospitalization, total health care cost and so on; 2) if $K > 1$ (heterogeneity), estimate a treatment effect for each group and compare the results across the groups; 3) if the effects vary over the groups, heterogeneity can be translated by each group's health profile (e.g. higher effectiveness found in sick but less hospitalized group). This approach is relatively conservative and combines multiple proxies objectively. Estimating K implies a near consensus of model selection criteria such as Bayesian Information Criteria (BIC), adjusted BIC, Akaike Information Criteria (AIC), and consistent AIC; and bootstrap likelihood ratio test (BLRT). Furthermore, it is difficult to find a practically useful K (say < 5) because K tends to diverge to N (i.e. each subject is a group), for a large enough sample size N. Applying heterogeneity estimation to a claims data of 3260 subjects for two years found two heterogeneous groups (BIC, adjusted BIC, consistent AIC, and BLRT all supported $K = 2$ except AIC). One group (N = 2841) was significantly sicker than the other group (N = 419) in Year 1 (and in Year 2) at 5%: Charlson Comorbidity Index 3.91 vs 0.11 (4.49 vs 0.14); length of stay 0.87 vs 0.03 (1.04 vs 0); total cost \$10690 vs \$245 (\$11149 vs \$184).

PMC50

EVIDENCE AND VALUE: IMPACT ON DECISION MAKING—THE EVIDEM FRAMEWORK AND POTENTIAL APPLICATIONS

Goetghebeur MM, Wagner M, Khoury H, Levitt R, Erickson LJ, Rindress D

BioMedCom Consultants Inc, Montreal, QC, Canada

Develop a quantitative and practical methodology to structure, objectify and facilitate health care decisionmaking. A conceptual framework was developed that segregated components of decision-making into three categories: 1) quality of evidence available; 2) intrinsic value of the health care intervention; and 3) extrinsic or system related value, usually not directly quantifiable. Using this framework, practical tools to assess health care interventions were designed drawing on an extensive review of the literature and of current decisionmaking processes for drug reimbursement around the world. A matrix to quantify the quality of evidence available for a health care intervention was